



Le Comité de surveillance Statistique

Délibération STAT n° 20/2017 du 10 octobre 2017

Objet : Demande formulée par l'Institut scientifique de santé publique afin de recevoir de la Direction générale Statistique – Statistics Belgium la communication de données d'étude codées dans le cadre du couplage des données d'échantillon de l'Enquête de santé 2013 avec des données du recensement administratif 2011 (STAT-MA-2017-020)

Le Comité de surveillance statistique (ci-après "le Comité") ;

Vu la loi du 4 juillet 1962 *relative à la statistique publique* (ci-après "la loi statistique publique") ;

Vu la loi du 8 décembre 1992 *relative à la protection de la vie privée à l'égard des traitements de données à caractère personnel* (ci-après "la LVP") ;

Vu l'arrêté royal du 13 février 2001 *portant exécution de la loi du 8 décembre 1992 relative à la protection de la vie privée à l'égard des traitements de données à caractère personnel* (ci-après "l'arrêté royal du 13 février 2001") ;

Vu l'arrêté royal du 7 juin 2007 *fixant les modalités relatives à la composition et au fonctionnement du Comité de surveillance Statistique* ;

Vu la demande de l'Institut scientifique de Santé publique, reçue le 18 juillet 2017 ;

Vu la demande d'avis technique et juridique adressée au Service public fédéral Économie, PME, Classes moyennes et Énergie (Direction générale Statistique - Statistics Belgium) en date du 10 septembre 2017 ;

Vu l'avis technique et juridique reçu le 5 octobre 2017 ;

Vu le rapport du Président ;

Émet, après délibération, la décision suivante, le 10 octobre 2017 :

I. OBJET DE LA DEMANDE

1. La demande vise à ce que l'Institut scientifique de Santé publique, ci-après le Chercheur, soit autorisé à recevoir de la Direction générale Statistique – Statistics Belgium (ci-après la DGSSB) la communication de données d'étude codées, et plus particulièrement :

- des données provenant de l'enquête de santé 2013 ;
- des données du recensement administratif 2011.

2. La demande vise également à ce que soit approuvé le contrat de confidentialité à conclure entre la DGSSB et le Chercheur à la suite de cette communication.

II. REMARQUE PRÉALABLE

3. Le Chercheur a déjà été autorisé précédemment à recevoir la communication de données provenant de l'enquête de santé dans le cadre d'un autre projet de recherche. Voir la décision suivante du Comité : délibération STAT n° 08/2016 du 10 mai 2016.¹.

III. EXAMEN DE LA DEMANDE

A. LÉGISLATION APPLICABLE

A.1. Loi statistique publique

4. Sur la base des articles 15 et 15*bis* de la loi statistique publique, la DGSSB est habilitée, après autorisation du Comité de surveillance statistique et moyennant un contrat de confidentialité approuvé

1

https://www.privacycommission.be/sites/privacycommission/files/documents/d%C3%A9lib%C3%A9ration_STAT_08_2016.pdf

par ce même Comité, à communiquer des données d'étude codées aux destinataires mentionnés dans la loi statistique publique aux conditions fixées dans cette même loi.

A.2. LVP et arrêté royal du 13 février 2001

5. En vertu de l'article 1, § 1 de la LVP et de l'article 1, 3° de l'arrêté royal du 13 février 2001, des données d'étude codées relatives à des personnes physiques identifiées ou identifiables constituent des données à caractère personnel dont le traitement n'est autorisé qu'aux conditions fixées dans la LVP et dans l'arrêté royal du 13 février 2001.

B. BASE JURIDIQUE

6. L'Institut scientifique de Santé publique fait partie des destinataires de données énumérés dans la loi statistique publique, au sens de l'article 15, premier alinéa, 1° de la loi statistique publique.

7. Le Chercheur entre donc en principe en ligne de compte pour être autorisé à recevoir les données demandées.

C. FINALITÉ

8. Les données à caractère personnel doivent être collectées pour des finalités déterminées, explicites et légitimes (article 4, § 1, 2° de la LVP).

9. En vertu de l'article 15 de la loi statistique publique, les données à caractère personnel codées doivent être obtenues pour poursuivre une finalité d'étude spécifique.

10. Dans la demande, il est question d'une utilisation à des fins d'étude scientifique et statistique. Le Chercheur mentionne trois finalités :

- Analyse de l'homogénéité des clusters en termes socio-économiques ; comme mentionné, les clusters (chaque fois 4 ménages) sont créés sur la base du secteur statistique, du groupe d'âge de la personne de référence et de la taille du ménage. La question est de savoir si ces ménages sont aussi similaires en termes de niveau de formation, de statut professionnel et de situation sur le marché du travail de la personne de référence (et du conjoint, s'il y en a un).
- Analyse des effets de substitution en ce qui concerne la composition – en termes socio-économiques – de l'échantillon finalement constitué par rapport à l'échantillon initial (c'est-à-dire avant l'application de la substitution). Concrètement, il s'agit (par exemple) de la question

de savoir si l'application de substitution implique que des personnes/ménages moins formés sont remplacés (substitués) par des personnes/ménages plus formés ou si l'application de la substitution est "neutre", c'est-à-dire qu'elle n'a pas d'impact sur la composition - en termes socio-économiques – de l'échantillon constitué.

- Ajout, aux informations utilisées pour la composition de l'échantillon, d'informations limitées (4 variables) provenant du Recensement administratif 2011.

11. L'Enquête de santé a pour but de dresser une représentation fiable de la santé de la population. Étant donné que le statut socio-économique n'a pas été utilisé comme critère pour composer l'échantillon, il est possible que le statut socio-économique des participants à l'enquête diffère par rapport à celui des personnes/ménages qui avaient été sélectionnées initialement pour participer à l'enquête. Le Chercheur veut dès lors réaliser deux finalités :

- évaluer l'homogénéité des clusters établis dans le cadre de l'Enquête de santé 2013 (tous les ménages sélectionnés pour participer à l'enquête)
- évaluer le processus de substitution en ce qui concerne les inégalités socio-économiques dans la participation à l'Enquête de santé (tous les ménages qui ont été contactés pour participer à l'enquête).

12. Concrètement, il s'agit de compléter/coupler les informations utilisées pour la composition de l'échantillon avec des informations limitées (4 variables) provenant du recensement administratif 2011. Cet ajout/ce couplage permet de réaliser les finalités poursuivies, étant donné que les informations sur le statut socio-économique des ménages qui sont invités à participer à l'enquête de santé ne sont pas ou quasiment pas présentes dans le Registre national qui est utilisé comme cadre pour l'échantillon en vue de l'Enquête de santé.

13. Les finalités de l'utilisation des données demandées sont déterminées, explicites et légitimes au sens de l'article 4, § 1, 2° de la LVP.

14. En vertu de l'article 15 de la loi statistique publique, les données à caractère personnel codées doivent être collectées à des fins statistiques ou scientifiques. D'après l'institution de gestion, la finalité des analyses que le Chercheur envisage de réaliser sur la base des données demandées est évidente d'un point de vue statistique. Il ne fait notamment aucun doute que la recherche menée par le Chercheur est d'une grande qualité et revêt une importance considérable, tant pour la recherche scientifique qu'en tant qu'outil dans le cadre de l'élaboration de politiques en matière de santé publique (on renvoie à cet égard à ce qui a déjà été réalisé précédemment² sur la base des enquêtes de santé

² Voir la note de bas de page n° 3 ci-avant.

de 2004, 2008 et 2013). Le Comité adhère à cet avis et confirme que les exigences en matière de finalité figurant dans la loi statistique publique sont respectées.

D. DONNÉES

15. Les données individuelles suivantes des enquêtes respectives sont demandées en vue des finalités d'étude précitées :

1. Population de l'échantillon Enquête de santé 2013

Les variables suivantes sont demandées :

- Le "ID_IND". Il s'agit du code d'identification de chaque membre des ménages de l'échantillon (il est composé du numéro de groupe/du numéro de trimestre, du numéro de cluster/du numéro d'ordre du ménage dans le cluster, et du numéro de l'individu au sein du ménage) ;
- SEC102011. Il s'agit du code pour le secteur statistique ;
- Npers. Il s'agit de la taille du ménage ;
- Agecat1. Il s'agit du groupe d'âge.

2. Les données du Recensement administratif 2011

Les variables suivantes sont demandées :

- CAS (la situation sur le marché du travail) ;
- EDU (niveau de formation / Eurostat)
- SIE (statut professionnel)
- CD_ISCED (niveau de formation codes ISCED)

16. Les données sont individuelles et ne concernent que les personnes faisant partie des ménages ayant participé à l'enquête. Elles comprennent les informations obtenues via le questionnaire écrit, le questionnaire en face à face et le questionnaire des ménages. Les informations relatives à la résidence du ménage sont limitées à l'arrondissement.

17. Les données précitées sont couplées et codées par la DGSSB.

E. PROPORTIONNALITÉ

E.1. Quant à la nécessité d'obtenir des données codées

18. Le Chercheur ne peut recevoir les données à caractère personnel codées que si un traitement de données anonymes ne permettait pas de réaliser les finalités statistiques ou scientifiques visées (article 4 de la LVP).

19. L'étude et l'analyse qui sont ici visées ne peuvent pas supporter la "perte de données" d'une éventuelle anonymisation par la DGSSB (par exemple, en reprenant les données demandées dans des tableaux indiquant des totaux).

20. La proposition d'étude prévoit un couplage de données d'échantillon et de données de terrain réalisé dans le cadre de l'Enquête de santé 2013 avec un nombre limité de données (socio-économiques) provenant du Recensement administratif 2011.

21. Le Chercheur a dès lors besoin des données à caractère personnel codées qui sont demandées. Une communication d'informations purement anonymes ne peut ici suffire.

22. Les finalités justifient donc le traitement de données à caractère personnel codées.

E.2. Quant à la quantité de données

23. Dans la demande, la proportionnalité est suffisamment démontrée, par finalité, pour toutes les données demandées.

24. L'ensemble de données qui sera communiqué pour l'étude est donc pertinent et dès lors adéquat et non excessif au sens de l'article 4, § 1, 3° de la LVP.

E.3. Quant au délai de conservation des données

25. Les données à caractère personnel ne peuvent pas être conservées pour une durée excédant celle nécessaire à la réalisation des finalités pour lesquelles elles sont obtenues et pour lesquelles elles sont traitées ultérieurement (article 4, § 1, 5° de la LVP).

26. Le Chercheur envisage de conserver les données pendant 18 mois. Le Chercheur estime ce délai nécessaire pour répondre aux éventuelles remarques formulées sur le manuscrit scientifique (adaptation éventuelle de l'analyse appliquée). Le Comité en prend acte.

F. DÉCLARATION

27. Avant de procéder à un ou à plusieurs traitements, automatisés en tout ou en partie, des données codées demandées en vue de réaliser les finalités envisagées, le Chercheur doit en faire la déclaration auprès de la Commission de la protection de la vie privée.

G. SÉCURITÉ

G.1. Conseiller en sécurité de l'information

28. D'après les documents transmis par le Chercheur, il apparaît que ce dernier dispose d'un conseiller en sécurité de l'information dont l'identité a également été communiquée.

G.2. Politique de sécurité

29. Il ressort des documents transmis par le Chercheur que ce dernier dispose d'une politique de sécurité.

30. D'après la déclaration de conformité accompagnant la demande de communication des données et le contrat de confidentialité, on peut établir que les 14 mesures en matière de sécurité généralement recommandées par la Commission de la protection de la vie privée lors du traitement de données à caractère personnel sont toutes réalisées par le Chercheur.

31. Le Chercheur a également été informé de ses obligations en matière de respect de la confidentialité. Un avis positif est donc en l'occurrence justifié, d'après l'institution de gestion.

G.3. Personne physique responsable

32. L'identité de la personne physique responsable a été communiquée. Celle-ci est personnellement responsable du respect de toutes les obligations concernant l'exécution de la loi statistique publique, de la LVP, de leurs arrêtés d'exécution, de toute autre disposition légale ou réglementaire de protection de la vie privée, des dispositions de la présente décision du Comité et des dispositions du contrat de confidentialité.

33. Cette personne exercera un contrôle effectif de l'utilisation licite des données fournies.

34. Les mesures dont il est question aux points G. 1. à G.3. inclus, qui doivent garantir la protection et la sécurité des données d'étude transmises, comme l'exigent l'article 16 de la LVP et

l'article 15*bis* de la loi statistique publique, sont efficaces si elles sont contrôlées et suivies de façon stricte dans la pratique.

G.4. Séparation des autres traitements

35. Le Chercheur doit séparer le présent traitement des données dont il est question ici pour les finalités indiquées des autres traitements de données à caractère personnel qu'il contrôle éventuellement.

G.5. Interdiction de décodage

36. Le Chercheur doit s'engager contractuellement à mettre en œuvre tous les moyens possibles afin d'éviter que soit retrouvée l'identité des personnes auxquelles les données à caractère personnel communiquées se rapportent.

G.6. Interdiction de couplage

37. Le Chercheur ne peut pas tenter de coupler les données à caractère personnel obtenues à des données à caractère personnel qui lui ont déjà été transmises en application d'autres autorisations.

G.8. Confidentialité

38. Le Chercheur s'engage à respecter la confidentialité des données d'étude et à veiller à ce que celles-ci ne soient utilisées que par les membres de son propre personnel en vue de l'exécution de la recherche visée.

G.9. Personnes qui utilisent les données d'étude et liste de ces personnes

39. Les personnes habilitées (deux personnes citées nominativement d'après la demande) qui utiliseront les données d'étude sont indiquées. Si d'autres chercheurs doivent travailler avec les données, le Comité en sera informé. Le Comité en prend acte et souhaite que le Chercheur le communique aussi à l'institution de gestion, le cas échéant.

40. Plusieurs données de l'ensemble de données demandé peuvent être considérées comme des données à caractère personnel au sens de l'article 7 de la LVP.

41. Vu le caractère sensible de ces données, le Chercheur doit respecter le chapitre III de l'arrêté royal du 13 février 2001, soit :

- établir une liste sur laquelle figurent les (catégories de) personnes qui utiliseront les données communiquées. Cette liste sera continuellement actualisée et tenue à la disposition du Comité et de l'institution de gestion ;

- ces personnes doivent être tenues au respect du caractère confidentiel des données visées par une obligation légale ou statutaire ou par une disposition contractuelle équivalente. Ces personnes signeront au moins une déclaration par laquelle elles s'engagent à préserver la sécurité et le caractère confidentiel des informations.

- la déclaration du traitement automatisé des données obtenues mentionne la loi ou le règlement autorisant le traitement de telles données à caractère personnel.

H. AUTRES CONDITIONS D'UTILISATION

H.1. Diffusion des résultats

42. Les résultats seront présentés de manière agrégée. On évitera ainsi que des données soient identifiées après publication des résultats.

43. Le Chercheur doit en effet veiller à ce qu'après l'analyse et l'utilisation des données, les résultats publiés restent anonymes et globaux de sorte que les données individuelles ne puissent pas être identifiées directement ou indirectement à partir de ces résultats.

44. Les résultats ne peuvent donc être diffusés que sous une forme globale et anonyme.

45. Au minimum deux semaines avant la diffusion, le Chercheur doit d'ailleurs soumettre la publication envisagée à la DGSSB.

H.2. Finalité scientifique, normes scientifiques et méthodes d'analyse

46. Le Chercheur n'est pas un destinataire au sens de l'article 15, premier alinéa, 4° de la loi statistique publique, mais un destinataire au sens de l'article 15, premier alinéa, 1° de la loi statistique publique. L'exigence de décrire les méthodes d'analyse qui seront utilisées dans le cadre de la recherche et celle d'apporter la preuve que le projet de recherche répond aux normes scientifiques en vigueur ne se posent en fait pas. Le Chercheur communique néanmoins ce qui suit à ce sujet.

47. Le Chercheur travaille dans le cadre d'un système de qualité. Pour les projets relatifs à une recherche épidémiologique, le Chercheur dispose d'un certificat ISO 9001. Les résultats de l'étude sont publiés dans des revues scientifiques, soumises à un peer review indépendant. En ce qui concerne les méthodes d'analyse, le Chercheur mentionne ce qui suit : "Les analyses envisagées sont de nature plutôt descriptive : il s'agit surtout de décrire la composition de l'échantillon initial et les glissements qui interviennent à la suite de la substitution. La régression logistique multiple sera appliquée afin d'évaluer une éventuelle distorsion (d'une sélection) des résultats de recherche (résultats de l'Enquête de santé).

48. D'après l'institution de gestion, les méthodes et la technique d'analyse envisagées répondent aux normes scientifiques. Pour autant que le Comité ait pu en juger, ni le caractère scientifique des finalités et des méthodes du projet de recherche, ni leur valeur scientifique ne sont contestables.

H.3. Contrôle

49. Le Chercheur accepte expressément que des représentants du Comité aient, à chaque instant et sans mise en demeure préalable, accès aux locaux et à l'infrastructure informatique où les données communiquées sont conservées, pour contrôler l'exécution des dispositions de la décision du Comité, des dispositions de la loi statistique publique et de ses arrêtés d'exécution ainsi que des dispositions du contrat de confidentialité.

50. Sur simple demande, le Comité peut obtenir l'accès à d'autres locaux et à d'autres systèmes ICT afin de contrôler si aucune violation des dispositions de sa décision, des dispositions de la loi statistique publique et de ses arrêtés d'exécution et des dispositions du contrat de confidentialité n'est commise.

I. LE CONTRAT DE CONFIDENTIALITÉ

51. Les données d'étude sont communiquées au Chercheur en vertu d'un contrat de confidentialité conclu entre la DGSSB et le Chercheur.

52. Le contrat de confidentialité, joint en annexe à la demande de communication des données, fixe les conditions auxquelles les données peuvent être transmises par la DGSSB et utilisées par le Chercheur.

53. Le contrat de confidentialité contient au moins les mentions légalement obligatoires telles que définies à l'article 15*bis* de la loi statistique publique, dont la durée du contrat de confidentialité, en l'occurrence jusque fin 2026 (collecte de données jusque fin 2023, analyse et rapport jusque fin 2026).

Cela ne signifie aucunement qu'au terme de ce délai contractuel, la confidentialité des données elles-mêmes puisse être rompue. Elle doit dès lors être respectée de manière illimitée dans le temps.

54. Les dispositions contractuelles relatives à la vie privée et à la confidentialité figurant dans le contrat de confidentialité sont reprises dans la présente décision du Comité, ce qui permet ainsi également à des personnes étrangères au contrat de confidentialité de s'adresser au Comité qui peut dans ce cas contrôler le respect des conditions auxquelles les données peuvent être utilisées par le Chercheur.

IV. DÉCISION GÉNÉRALE

55. Lors du traitement des données à caractère personnel obtenues, le Chercheur doit tenir compte de la LVP, de la loi statistique publique, de leurs arrêtés d'exécution et de toute autre disposition légale ou réglementaire de protection de la vie privée, des dispositions de la présente décision du Comité et des dispositions du contrat de confidentialité qu'il aura conclu avec la DGSSB.

V. DÉCISION SPÉCIFIQUE

58. Le Comité estime que :

- le Chercheur dispose d'un fondement légal pour réclamer les données d'étude codées demandées ;
- la communication par la DGSSB au Chercheur des données d'étude codées demandées est autorisée en vue des finalités visées, selon les délais, modalités et limitations prévus dans la présente délibération ;
- la durée de conservation des données et donc la durée du contrat de confidentialité sont limitées à 6 mois, à compter de la fin de la recherche qui durera maximum 12 mois. Au terme du délai de conservation, la confidentialité des données elles-mêmes doit être respectée de manière illimitée dans le temps ;
- vu la nature de certaines données de l'ensemble de données demandé, le Chercheur doit respecter le chapitre III de l'arrêté royal du 13 février 2001 ;
- après la recherche, les données ne peuvent être publiées qu'à un niveau suffisamment agrégé.

PAR CES MOTIFS,

Le Comité

autorise la DGSSB à communiquer à l'Institut Scientifique de la Santé Publique les données à caractère personnel susmentionnées, aux conditions précitées ;

approuve le contrat de confidentialité y afférent, aux conditions précitées.

L'Administrateur f.f.,

Le Président,

(sé) An Machtens

(sé) Gert Vermeulen